

Análisis y modelado digital de la voz

Técnicas recientes y aplicaciones



Hugo L. Rufiner

UNIVERSIDAD
NACIONAL DEL LITORAL



Análisis y modelado digital de la voz. Técnicas recientes y aplicaciones

Digital analysis and modeling of speech:
Recent Techniques and applications

Consejo Asesor
Colección Ciencia y Técnica

Rufiner, Hugo Leonardo.

Análisis y modelado digital de la voz. Técnicas recientes y aplicaciones. - 1a ed. -

Santa Fe: Universidad Nacional del Litoral, 2009.

284 pp.; 25 x 17 cm. (Ciencia y técnica)

ISBN 978-987-657-090-9

1. Matemática aplicada. 2. Teorías de la información I. Título

CDD 530.15

Coordinación editorial: *Ivana Tosti*

Corrección: *Laura Prati*

Diseño original de interiores: *José Díaz*

Diagramación de interiores y diseño de tapa: **mzD** mágico zoo Diagonal)

© Hugo Leonardo Rufiner, 2009.



© edicionesUNL

Secretaría de Extensión,
Universidad Nacional del Litoral,
Santa Fe, Argentina, 2009.

Queda hecho el depósito que marca la ley 11.723.

Reservados todos los derechos.

9 de julio 3563 (3000)

Santa Fe, Argentina.

Telefax: (0342) 4571194

editorial@unl.edu.ar

www.unl.edu.ar/editorial

Aplicaciones y modelado digital de la voz. Técnicas recientes y aplicaciones

Digital analysis and modeling of speech:
Recent Techniques and applications

Hugo Leonardo Rufiner

Prefacio

La emulación de la comunicación humana en forma artificial ha sido una meta largamente perseguida. Alcanzarla permitiría interactuar con las máquinas de una manera más sencilla y completamente distinta a la actual. A pesar de la aparente sencillez con la que los humanos realizan esta tarea, y de más de tres décadas de investigación, su resolución completa por medios artificiales ha permanecido fuera del alcance. Los resultados presentados en este libro son parte de los obtenidos en la tesis de Doctorado del autor [138] y constituyen un aporte en la dirección de esta meta. Mediante un enfoque interdisciplinario se explora la idea acerca de cómo definir y buscar una representación "óptima" de la señal de voz mediante técnicas no convencionales, basadas en la utilización de diccionarios discretos, a los fines de su posterior manipulación mediante un sistema artificial.

Para llegar a presentar este libro se ha recorrido un camino de investigación de varios años. Durante la tesis de grado se encaró el estudio inicial de los sistemas de *reconocimiento automático del habla* [139]. Se diseñó y probó un sistema que utilizó la *transformada de Fourier* para la etapa de procesamiento, *redes neuronales* para la acústico-fonética y *sistemas expertos* para el análisis del lenguaje.

Posteriormente se empleó un enfoque más biológico para la etapa de procesamiento basado en *modelos de oído*, y se inició también el estudio de la aplicación de los *modelos ocultos de Markov* al problema del reconocimiento

[136]. Aquí comenzó a surgir la idea de comparar las técnicas clásicas, sustentadas generalmente en estimadores espectrales, con un análisis más similar al que realiza el sistema auditivo.

En la tesis de Maestría se inició la comparación entre el análisis basado en la denominada *transformada ondita* y el basado en la clásica transformada de Fourier [137]. Se exploró principalmente el caso de la *transformada ondita discreta diádica*, en contraste con la *transformada de Fourier de corta duración*. A pesar de sus atractivas cualidades para el procesamiento de señales transitorias, la comparación resultó desfavorable para este tipo particular de transformación basada en onditas. Allí se advirtió la multiplicidad de facetas que presentaban las representaciones relacionadas con onditas y la dificultad de brindar una respuesta definitiva acerca de cuál resultaba mejor en este contexto.

Comenzando el trabajo para la tesis de Doctorado [138] se continuó la exploración de la aplicación de onditas, tratando de superar algunos de los problemas detectados previamente, y ello resultó en mejoras significativas [45, 170, 140, 171, 143, 172, 160]. Dentro de los caminos alternativos explorados aparecen principalmente las representaciones ralas y/o independientes [141, 145, 144] y la inclusión de información adicional en la representación [146].

A partir de una recomendación del jurado de tesis, el material disponible originalmente se ha reorganizado con el fin de servir como base de revisión y actualización para docentes o investigadores de áreas relacionadas con el análisis del habla; como por ejemplo profesionales que trabajan en procesamiento de señales, matemática aplicada o lingüística. El libro supone conocimientos básicos de *análisis funcional*, *análisis de señales* y *teoría de la información* que pueden encontrarse en algunos de los textos clásicos de cada área (como por ejemplo [6], [132] y [27] respectivamente, o en forma resumida en [138]).

Se puede decir que este trabajo aporta una visión alternativa al problema del análisis y la representación de una señal a partir de un *modelo* más “complejo” que el tradicional. En este sentido sigue una tendencia presente en la ciencia moderna que implica prácticamente un cambio de paradigma. Se ha cambiado el enfoque de la *simplicidad*, que intentaba analizar la realidad mediante su descomposición en fenómenos sencillos,

por el de la *complejidad*, que intenta incluir muchos más aspectos de la realidad desde la concepción inicial del modelo [117].

Organización del libro

El libro se organiza de la siguiente forma: en el Capítulo 1 se presenta la introducción al trabajo, destacando la motivación inicial y los conceptos fundamentales que guían el desarrollo del mismo. Estas ideas se ilustran con algunas notas históricas relevantes.

Las bases fisiológicas de la comunicación humana se presentan en el Capítulo 2. Se exponen aquí las características principales de la señal de voz, del aparato fonador y del sistema auditivo. El énfasis está puesto en los aspectos neurosensoriales de este último que permiten identificar las pistas acústicas del habla.

En el Capítulo 3 se exponen los fundamentos matemáticos del *análisis de señales*. Se discuten las diferentes formas de representación de una señal, desde las tradicionales hasta las más recientes o *no convencionales*, incluyendo aquellas más orientadas al caso de la señal de voz. El enfoque está orientado principalmente a señales de tiempo continuo.

Una revisión del modelado de señales, que sirve como marco conceptual de varias de las técnicas utilizadas, se desarrolla en el Capítulo 4.

En el Capítulo 5 se contemplan diferentes alternativas para obtener la representación de señales de tiempo discreto a partir de *diccionarios*, en particular los casos ortogonales basados en Fourier, onditas, paquetes de onditas y paquetes de cosenos.

Los fundamentos de las *representaciones ralas y/o independientes* obtenidas a partir de diccionarios se presentan en el Capítulo 6, y se mencionan sus ventajas y desventajas, las conexiones con otros enfoques y algunos ejemplos de aplicación.

En el Capítulo 7 se desarrollan diferentes métodos basados en técnicas no convencionales y se estudia su aplicación al análisis y la representación de la señal de voz. Se comparan las alternativas propuestas con las clásicas, desde aspectos cualitativos, medidas de eficacia y resultados de aplicación a problemas de clasificación de fonemas, reconocimiento del habla y limpieza de ruido. Finalmente, en el Capítulo 8 se discuten las conclusiones generales junto con los posibles trabajos futuros.

Índice general

Glosario	5
Prefacio	9
1. Introducción	12
1.1. Motivación	12
1.2. Objetivos	17
1.3. Técnicas convencionales	18
1.3.1. Análisis de señales	18
1.3.2. Análisis de Fourier	19
1.3.3. Bases y transformaciones lineales	20
1.3.4. Análisis de Gabor	20
1.3.5. Distribuciones tiempo-frecuencia	22
1.3.6. Análisis específicos para el habla	22
1.4. Técnicas no convencionales	23
1.4.1. Análisis basado en onditas	23
1.4.2. Análisis ralo y/o independiente	25
1.5. Comentarios de cierre del capítulo	27
2. Bases fisiológicas de la comunicación	30
2.1. Introducción	30
2.2. Mecanismo de producción del habla	33
2.2.1. Aparato fonador	36
2.2.2. Sonidos y fonemas	40
2.2.3. Segmentos, suprasegmentos y sílabas	43
2.3. Señal de voz	43
2.4. Fisiología de la audición	48
2.4.1. Recepción y adecuación acústica	48
2.4.2. Transducción mecánico-eléctrica	49

2.4.3.	Nervio auditivo y codificación nerviosa	53
2.4.4.	Vía auditiva	59
2.4.5.	Corteza auditiva	61
2.5.	Percepción	65
2.5.1.	Inteligibilidad	67
2.5.2.	Algunos experimentos perceptuales	67
2.6.	Comunicación en condiciones adversas	69
2.6.1.	Ruido y reverberación	69
2.6.2.	Humanos y máquinas	69
2.7.	Comentarios de cierre del capítulo	72
3.	Análisis y representación de señales	74
3.1.	Introducción	74
3.2.	Análisis lineal invariante en el tiempo	76
3.2.1.	Transformada de Fourier	76
3.3.	Análisis lineal no estacionario	77
3.3.1.	Señales analíticas y frecuencia instantánea	81
3.3.2.	Transformada de Fourier de corta duración	82
3.3.3.	Transformada ondita	85
3.4.	Análisis no lineal y/o no estacionario	88
3.4.1.	Distribuciones $t - f$ cuadráticas	89
3.4.2.	Representaciones $t - f$ no lineales	92
3.5.	Análisis específicos para el habla	95
3.5.1.	Coefficientes de predicción lineal	95
3.5.2.	Análisis cepstral	97
3.5.3.	Análisis predictivo lineal perceptual	100
3.5.4.	Modelos auditivos	101
3.6.	Aspectos relacionados con la robustez	103
3.7.	Comentarios de cierre del capítulo	104
4.	Modelización de señales	106
4.1.	Introducción	106
4.2.	Pautas para evaluar un modelo	107
4.2.1.	Propósito de este trabajo	108
4.2.2.	Redundancia e independencia estadística	109
4.2.3.	Códigos locales, completamente distribuidos y malos	110
4.2.4.	El problema del ruido	112
4.2.5.	Otros aspectos a considerar	112
4.3.	Medidas de "calidad"	114
4.3.1.	Medidas estadísticas y de información	115
4.3.2.	Medidas de dispersión	117
4.3.3.	Medidas de ajuste	119
4.4.	Comentarios de cierre del capítulo	120
5.	Representaciones basadas en diccionarios discretos	122
5.1.	Introducción	122

5.2.	Transformada discreta de Fourier	123
5.2.1.	Transformada discreta de Fourier de corta duración	125
5.3.	Transformada ondita discreta	128
5.3.1.	Transformada ondita rápida	132
5.3.2.	Familias de onditas	136
5.4.	Transformada paquetes de onditas	139
5.4.1.	Transformada paquetes de ondita rápida	143
5.5.	Transformada paquetes de cosenos	144
5.6.	Comentarios de cierre del capítulo	145
6.	Representaciones ralas y/o independientes	150
6.1.	Introducción	150
6.2.	Ventajas y desventajas	153
6.3.	Planteo del problema	155
6.4.	Selección de coeficientes o inferencia	159
6.4.1.	Caso limpio, enfoque determinístico	159
6.4.2.	Caso ruidoso	165
6.5.	Búsqueda del diccionario o aprendizaje	173
6.5.1.	Diseño a medida	173
6.5.2.	Ajuste automático	174
6.6.	Comentarios de cierre del capítulo	176
7.	Aplicaciones a la señal de voz	178
7.1.	Introducción	178
7.2.	Descripción de los experimentos	179
7.3.	Representaciones convencionales	181
7.4.	Inclusión de cambios de complejidad	185
7.5.	Representaciones basadas en onditas	188
7.5.1.	Transformada discreta diádica	188
7.5.2.	Transformada paquetes de onditas	192
7.6.	Representaciones ralas y/o independientes	201
7.6.1.	Diccionarios a medida	201
7.6.2.	Diccionarios óptimos	222
7.7.	Comentarios de cierre del capítulo	253
8.	Consideraciones finales	256
8.1.	Conclusiones	256
8.2.	Perspectivas futuras	257
	Bibliografía	268

Se diagramó y se compuso en  edicionesUNL y se terminó
de imprimir en Artes Gráficas Papiros SACI, Castro Barros 1395/97,
Buenos Aires, Argentina, junio de 2009.